

MUSICAL SOUND CODING METHOD, MUSICAL SOUND DECODING METHOD AND RECORDING MEDIUM RECORDED WITH PROGRAM TO EXECUTE THESE METHODS

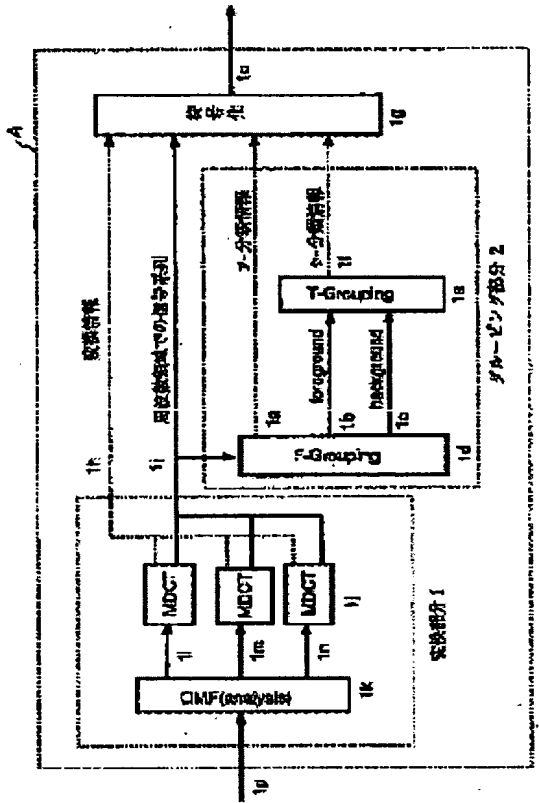
Patent number: JP2001318684
Publication date: 2001-11-16
Inventor: CHIKIRA KAZUAKI; IWAGAMI NAOKI
Applicant: NIPPON TELEGRAPH & TELEPHONE
Classification:
- International: *G10L11/00; G10L19/00; G10L19/02; H03M7/30; G10L11/00; G10L19/00; H03M7/30; (IPC1-7): G10L11/00; G10L19/00; G10L19/02; H03M7/30*
- european:
Application number: JP20000134625 20000508
Priority number(s): JP20000134625 20000508

Report a data error here

Abstract of JP2001318684

PROBLEM TO BE SOLVED: To realize a coding and decoding method that eliminates the waste and easily edits and processes for the system in which acoustic information being transmitted is received, temporarily recorded in a recording medium, reproduced, edited, processed and speed varied resulting in a low transmission efficiency and unnecessary amount of computations. **SOLUTION:** Acoustic signals in a time domain are converted into a time frequency domain by a band dividing through QMF and MDCT and classification information of coefficient groups by the size of the absolute value of the conversion coefficients and time information of detecting the generation and the extinction of peaks in a frequency domain are coded along with the coefficient groups in the time frequency domain obtained above. By conducting the coding indicated above, efficient transmission of acoustic signals and editing and processing at the decoding side are easily executed.

図 1



Data supplied from the *esp@cenet* database - Worldwide

【特許請求の範囲】

【請求項1】音響信号を時間周波数領域における係数群に変換する第一段階と、該第一段階から入力された時間周波数領域上における音響信号の係数群を、少なくとも2つのグループに分類し、この分類情報を出力する第二段階とを有し、前記係数群と前記分類情報とを符号化信号の系列として出力することを特徴とする楽音符号化方法。

【請求項2】請求項1記載の楽音符号化方法において、前記第一段階においては、入力された音響信号を高域信号と低域信号に帯域分離し、低域信号をさらに低高域信号と低低域信号とに帯域分離し、前記高域、低高域、低低域の各信号を予め定められたサンプル数ごとにそれぞれ時間領域から時間周波数領域へ変換し、前記変換により得られた時間周波数領域での係数群と、前記変換処理を行うサンプル数単位の時間長に対応する窓長の情報とを出力することを特徴とする楽音符号化方法。

【請求項3】請求項1記載の楽音符号化方法において、前記第二段階の処理に際し、前記第一段階により得られた時間周波数領域の係数群を、絶対値の大小によって少なくとも2つのグループに分配することを特徴とする楽音符号化方法。

【請求項4】請求項1記載の楽音符号化方法において、前記変換により得られた係数群を、時刻ごとに区切ることによって、各グループへの分類を行うことを特徴とする楽音符号化方法。

【請求項5】請求項4記載の楽音符号化方法において、音響信号のエネルギーに急激な変化のある時刻および、多数のフレームにわたって持続している特定の周波数成分が発生または消滅の何れかとなる時刻を区切りの時刻として各グループへの分類を行うことを特徴とする楽音符号化方法。

【請求項6】請求項1記載の楽音符号化方法において、前記第二段階の処理に際し、前記第一段階により得られた時間周波数領域の係数群を絶対値の大小によって少なくとも2つのグループに分配し、さらに前記各グループの音響信号を時刻ごとに区切ることによって、前記各グループへの分類を行うことを特徴とする楽音符号化方法。

【請求項7】音響信号を時間周波数領域における係数群に変換し、さらにこの変換された時間周波数領域上における音響信号の係数群を、少なくとも2つのグループに分類し、これら係数群と分類情報とを符号系列として出力する楽音符号化方法により得られた符号化信号の系列において、前記符号化信号の系列に含まれる前記分類情報を用いて前記係数群を構成する各時間周波数変換係数のグループを決定し、各グループに属する時間周波数変

に変換し、さらにこの変換された時間周波数領域上における音響信号の係数群を、少なくとも2つのグループに分類し、これら係数群と分類情報とを符号系列として出力する楽音符号化方法により得られた符号化信号の系列において、前記符号化信号の系列に含まれる前記分類情報を用いて前記係数群を構成する各時間周波数変換係数のグループを決定し、前記グループのうち特定グループに属する時間周波数変換係数のみを時間領域に変換して符号化信号の系列を再生する楽音復号化方法。

【請求項9】音響信号を時間周波数領域における係数群に変換し、さらにこの変換された時間周波数領域上における音響信号の係数群を、少なくとも2つのグループに分類し、これら係数群と分類情報を符号系列として出力する楽音符号化方法により得られた符号化信号の系列において、前記符号化信号の系列に含まれるサンプルごとに前記係数群を構成する各時間周波数変換係数から時間領域に変換して音響信号を再生することを特徴とする楽音復号化方法。

【請求項10】請求項1乃至請求項6の何れかに記載の楽音符号化方法を実行するプログラムを記録したことを特徴とする記録媒体。

【請求項11】請求項7乃至請求項9の何れかに記載の楽音復号化方法を実行するプログラムを記録したことを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】音響信号をデジタル信号として、編集加工、早送り再生、任意グループ再生等符号化出力を復号化するに際し、各種データ処理を容易にし、かつ利用し易い形式として表現する符号化および復号化方法に係る。

【0002】

【従来の技術】デジタル化された音響信号を、聴感上のひずみを抑えつつ、その信号を表現するビット数を削減するのが従来の主な符号化技術であった。その代表的な例としてTwinVQ、AAC、ATRAC（登録商標）等が挙げられる。一方、このような周波数帯域の圧縮を目的とした符号化方式とは異なり、符号としての構造に価値を持たせることを目的とした方式も普及している。これは音楽を構成する楽器音それぞれについて、ノートオン、ノートオフの時刻や音の高さ、強弱を符号として保持する方式であり、そのパラメータを変更することで容易に音楽の編集加工をすることができるとが最大の利点である。具体的にはMIDI（Musical Instrument Digital Interface）やStructured Audioなどの方式が挙げられるが、これらの方式の欠点はデジタルお

の音響信号からMIDI形式の符号へと自動変換する技術も研究されているが、この技術では音響信号が復号されたときに原音に近い音を再生することが困難である。

【0003】編集加工の一例として早送りが挙げられるが、音響信号を、上述のような構造化された符号に変換せず、直接編集し早送りを実現する技術が存在する。ここで早送りとは、基本的にピッチを変えずにテンポ（速さ）のみを早くすることを意味するものである。この技術は質の高い早送りを行うことができるが、従来の符号化技術と組み合わせて利用するためには一旦すべての音響信号を復号する必要がある、伝送経路を考慮に入れた場合、N倍の早送りをするためにはN倍の伝送容量が必要となる。また、この操作を実行する装置にかかる負荷は大きいことが問題となる。例えばTwinVQやCDによる方式ではフレームを飛ばして符号を伝送することによってこの問題を回避している。しかしながら早送りの質に問題があり、質と伝送効率を同時に満たす技術が求められている。

【0004】

【発明が解決しようとする課題】従来技術のように音響信号を圧縮する場合、復号された音響信号をそのまま再生するだけでなく、再び編集加工を施したり、音楽検索の対象として利用する場合等が考えられる。従来の方式では、そういった操作をする前に音響信号を送信し復号しなければならず、伝送効率や計算量の面で無駄が生じる。本発明は、符号の形式のまま編集加工が可能であり、符号化系列そのものが直接編集が可能のように構造化されており、意味をもつような符号化方法を提供することを課題としている。また上記のような構造化された符号系列を、入力音に近い音を再生できることを保証した上で、入力された音響信号から目的とする信号を自動的に生成する符号化および復号化方法およびこれら方法を実現する装置を提供することを課題とする。

【0005】

【課題を解決するための手段】上記課題を達成するために、本発明においては入力音を伝送再生するためのデジタル信号系列を保持すると同時に、それに加えて編集加工その他各種処理を行うための情報を符号系列に付加する。この情報は入力音をオブジェクトという塊に分離するための情報であり、この塊を単位として編集加工を可能とする。またこの情報は、後の編集加工の度に同様な操作を繰り返さないで済むように、符号化を行う段階で生成しておき、状況に応じて、必要な情報と信号系列とを符号系列から抽出し伝送する構成としている。すなわち、本発明においては基本的には、音響信号を時間周波数領域における係数群に変換する第一段階と、この第一段階から入力された時間周波数領域上における音響信号

て出力する楽音符号化方法としている。

【0006】ここで、前記楽音符号化方法に記載の第一段階において、入力された音響信号を高域信号と低域信号とに帯域分離し、低域信号をさらに低高域信号と低低域信号とに帯域分離し、前記高域、低高域、低低域の各信号を予め定められたサンプル数ごとにそれぞれ時間領域から時間周波数領域へ変換し、この変換により得られた時間周波数領域での係数群と、このMDCTの変換処理を行う際に用いられるサンプル単位の時間長に対応する窓長の情報とを出力する楽音符号化方式についても開示している。

【0007】前記楽音符号化方法の第二段階において、前記の時間周波数領域における係数群を、絶対値の大小によって少なくとも2つのグループに分配することについても開示している。

【0008】前記楽音符号化方法の第二段階において、前記係数群を、時刻ごとに区切ることによって、各グループへの分類を行うこと方法についても開示している。

【0009】また、前記の各グループへの分類を行う分類方法において、音響信号のエネルギーに急激な変化のある時刻および、多数のフレームにわたって持続している特定の周波数成分が発生あるいは消滅の何れかの状態となる時刻を区切りの時刻として各グループへの分類を行う方法についても開示している。

【0010】前記楽音符号化方法の第2段階において、絶対値の大小により分配する際、グループに分配された時間周波数係数群を時刻ごとに区切り、更に前記各グループへの分配を行う方法について開示している。

【0011】前記楽音符号化方法により得られた符号化信号の系列に含まれる前記分類情報を用いて前記係数群を構成する各時間周波数変換係数のグループを決定し、各グループに属する時間周波数変換係数の先頭部分をおのおの時刻順につなぎ合わせて時間領域に変換してもとの音響信号の系列を再生する複合方法について開示している。

【0012】前記符号系列に含まれる前記分類情報を用いて前記係数群を構成する各時間周波数変換係数のグループを決定し、前記グループのうち特定グループに属する時間周波数変換係数のみを時間領域に変換して信号系列を再生する復号化方法について開示している。

【0013】前記符号系列に含まれる前記サンプルごとに前記係数群を構成する各時間周波数変換係数から時間領域に変換して信号系列を再生することを特徴とする楽音復号化方法についても開示している。

【0014】上記の符号化方法及び復号化の処理において、これら符号化及び復号化処理を実行するプログラムを記録した記録媒体についても開示している。

の構成例を示したものである。また、図2は本発明における復号化方法の構成例を示したものである。まず図1の符号化方法について説明する。エンコーダAは変換部分1、グルーピング（分離）部分2、符号化部分1gの3つのモジュールから構成される。変換部分1のモジュールでは入力信号のanalysis（分解）用としてQMF（Quadrature Mirror Filter）1kが用いられている。このQMFは入力信号の帯域を2等分してサブバンドに分解するもので、本実施の形態においては、これを2段組み合わせることによって、入力された音響信号1pをフレームごとに、先ず第1段階で高域／低域に分解し、更に第2段階で低高域／低低域に分解し、これにより高域1l、低高域1m、低低域1nの3帯域に分割し、それぞれの信号が後続のMDCT（Modified Discrete Cosine Transform；離散コサイン変換）により処理される。ここで、入力信号の処理単位をフレームと定義し、1フレーム＝2048サンプルとする。なお、MDCTは時間領域の信号を或るサンプル数単位で周波数領域に変換する手法で、このサンプル数単位をこの変換における窓長と称している。

【0016】高域では時間変化の大きい信号が主であるため、短い窓長で信号を切り出すMDCTが適切である場合が多いが、一般に低域では時間変化が比較的小ないためその逆となる。本方式では、MDCTを行う前にQMF 1kで帯域分割することにより、このような窓長の問題を解決し、よりフレキシブルな変換を行なう。これにより得られた上記の各帯域に対応する3つの信号系列1l、1m、1nそれぞれに対して、MDCT 1jの処理を行う。ここで、MDCT 1jの窓長は可変であり、時間的変化の大きい信号に対しては短い窓長を使用し、比較的定常で時間的変化の少ない信号に対しては長い窓長を使用する。これによってMDCT処理後の信号のエネルギー分布を少ない係数の範囲に集中させ、その結果効率的に信号を符号化することが可能となる。ここで係数とは、例えば時間周波数領域において時間軸方向と周波数軸方向とで囲まれたデータ系列の単位領域を示すもので、この単位領域の集合を窓関数と呼んでいる。具体的には、予め用意された複数の窓関数（例え

ば本実施の形態における実験では32個、128個、512個の3種類用意した。）を利用して後に記述するグルーピングの操作までを行い、グループ間での音の偏りの度合いが高い窓長を選択する。この窓長選択基準となる音の偏りの度合いとは、あるグループに属する係数（上記単位領域）の数に対するそのグループに含まれる音響信号のエネルギーの比で求められる。すなわち、（グループ0のエネルギー）／（グループ0に属する係数の数）

が最大となる窓関数を選択する。ここで、グループ0は、絶対値の大きい係数群からなるグループである。このことは係数群の絶対値の大小によって複数のグループに分配し得ることを示している。窓長は、それぞれの周波数帯域で独立に決定する。図1に示すように、このようにして得られた各周波数領域ごとの信号系列1iは符号化部分となるモジュール1gに入力される。また、使用する窓長の情報を、変換情報1hとしてこの符号化部分のモジュール1gへと入力する。

【0017】グルーピング部分のモジュール2においては、変換部分のモジュール1から入力された各周波数ごとの信号系列1iを、周波数領域と時間領域とにグルーピングし、幾つかの信号の塊とするためのグルーピング情報としてF-分類情報1aおよびT-分類情報1fを生成する。これら2つの情報1aおよび1fを符号化部分のモジュール1gへと入力する。上記のグルーピングは、まず周波数領域で行い、つぎに時間領域で行う。周波数領域でのグルーピングは、図1においてF-Grouping 1dで示される部分で実行され、変換部分のモジュール1の出力した信号系列1iの係数をそれぞれの属するグループへ振り分ける操作である。振り分けは、ある閾値を設定し、信号系列1jの係数がその閾値よりも大きければ、foreground 1bのグループ、小さければbackground 1cのグループへ、という方法で行う。具体的には、信号系列を{f₀, f₁, f₂, …, f_{N-1}}から、グルーピング情報{g₀, g₁, g₂, …, g_{N-1}}を得ることである。ただし、

【数1】

$$g_i = \begin{cases} 0 & |f_i| < C \text{ のとき} \\ 1 & |f_i| \geq C \text{ のとき} \end{cases} \quad (\text{数1})$$

とし、Cは閾値である。ここで、0と1はグループ個々の値であり、この構成例は、2つのグループを持つ時の例であり、foregroundを1、backgroundのグループを0としている。

【0018】つぎに、得られた複数のグループの信号系

し、周波数領域信号1bおよび1cを時間信号に逆変換する。ここで、IMDCT 2cは前記DCTの逆変換を行うもので、周波数領域に変換されていたデータを時間領域のデータに復元するものであり、synthesis QMF 2dはサブバンドに分解されていた信号を

ば、周波数領域の foreground 成分 1 b において伝送される時間信号系列 $\{b_i\}$ は、周波数領域での

$$\{\alpha_i\}, \alpha_i = \begin{cases} f_i & g_i = 1 \quad \text{のとき} \\ 0 & g_i \neq 0 \quad \text{のとき} \end{cases} \quad (\text{数 } 2)$$

を時間領域に変換したものである。

【0019】 T-grouping 1 e における時間領域でのグルーピングは、時間軸方向で信号を切り分ける操作であり、具体的には、時間信号のエネルギーに急峻な変化のある時刻、あるいは「或る周波数成分」が発生および消滅した時刻を抽出することである。ここでのグルーピングは、人間が音を聴いたとき、音の区切りと判断する単位で音を塊とすることを目的としているので、上述の2種類の時刻をその判断基準として利用することによって、この目的に適した時刻を抽出する。この手順の一例を以下に示す。

【0020】 (操作1) 入力信号に対してFFT (Fast Fourier Transform) を行ない周波数領域でのピークを抽出する。このピークは聴感上で周波数弁別が可能となる最小の帯域幅に対応したパークスケールバンドごとに検索し、その条件は

1. (該当パークスケールバンド全体のエネルギー) / 5以上のエネルギーを有していること、
2. ひとつ隣りのパークスケールバンドの周波数成分よりも大きいエネルギーを有していること、
3. (ふたつ隣りの周波数成分) $\times 1.4$ よりも大きいエネルギーを有していること、

の3条件であり、これらの条件を満たすピークの周波数と周波数成分をすべて保持するものである。また、この操作1の入力信号として、各フレームごとの時間信号 2 0 4 8 サンプルを、1 2 8 サンプルずつ時間軸上を左にシフトさせ、シフト後の右側の空き領域には後続のサンプルを移動して埋め、これにより現在のフレームからFFTにより抽出されたピークそれぞれについて、ひとつ前の(操作1)で抽出されたピークと一致する周波数がないかどうか調べる。もし一致するピークがあるならば、周波数成分の増加分を変数 diff に加える。こうして得られた変数 diff が閾値である Onset レベルよりも大きければ、新たな音が出現したとして、この時刻を記録する。また、変数 diff の値が閾値である Offset レベルよりも小さくなった場合は、音が消えたとしてこの時刻を記録する。時刻はフレームと、そのフレーム内での時間軸上での上記シフト量で表現するが、以下ではそれらを単純に $\{T_i\}$ と表わしている。ここで検出された時刻の系列 $\{T_i\}$ を時間分離情報 (T-分離情報) 1 f として、符号化部分のモジュール 1 g に入力する。また、時刻の系列 $\{T_i\}$ は前述の時間信号

系列

【数 2】

領域のデータを所定の時間間隔で把握しグルーピングしていることになる。

【0021】 符号化部分のモジュール 1 g では、上述の3つの情報、すなわち変換情報 1 h、周波数分離情報 1 a および時間分離情報 1 f に加えて、変換部分が出力する信号の周波数領域での系列 1 i を、ビットストリーム 1 o に変換し符号化された信号とする。ビットストリーム 1 o の1例を図3および図4に示す。この例では一切のデータ圧縮を行っていないが、ヘッダにフレーム総数 N やフレーム長、またデータの圧縮方式などを記録することによって、さらなる可逆圧縮を施すことも可能である。以上述べた処理のアルゴリズムにしたがったプログラムを予め記録媒体に記録しておき、この記録媒体を用いて上記一連の処理の制御を実行させることにより、以上の符号化処理を容易に行うことが出来る。

【0022】 ここで、図2に示す復号系における復号化方法について説明する。本発明による符号化法では、符号化の時点で付加情報を抽出しているため、それらを利用して幾つかの機能を実現することが可能である。この実施の形態では、従来通りの信号復元、早送り再生および任意のグループのみを再生の3例について説明する。従来通りのデコーダとしての機能は、信号の周波数領域での信号系列 2 e と変換情報 2 f とだけを利用し、逆変換をすることによって実現する。これは、図1における符号化方法との逆の手順である。グループ選択、編集加工部 2 b において、2つの分離情報すなわち T-分類情報 2 g と F-分類情報 2 h を使用せず、周波数領域での信号系列 2 e を該当する3つの帯域成分 2 i、2 j、2 k にわけて3つの IMDCT 2 c にそれぞれ入力する。この IMDCT 2 c では MDCT の逆の処理を行い、それらの結果を QMF (synthesis) 2 d で合成する。これによって、ほぼ原信号通りの信号を再生することが可能となる。

【0023】 また、周波数分離情報 (F-分類情報) 2 h と時間分離情報 (T-分類情報) 2 g とを利用して、音響信号の早送り再生を簡単に行うことができる。たとえば、1 サンプル係数ごとに時間信号を破棄するといった方法で2倍の早送りをしようとする、音の周波数成分に変化が生じ、自然な早送りとはならない。そこで、従来の早送りの技術では音の構成単位となっている楽器音の発生から消滅までを一つの塊として、その後半を削除して再生していた。この場合、主旋律となっている楽

本発明では主旋律と背景音を分離してから時間軸上の音響信号の一部削除を行い、残りの音響信号を順次つなぎ合わせるにより自然な早送りとなる。具体的な手法は以下のとおりである。

【0024】まずグループ選択、編集加工部2bにおいて周波数分離情報(F-分類情報)2hをもとに、信号を複数の音響信号に分離する。それぞれの音響信号に対

$$\{b_k\}, \frac{(N-1)T_j + T_{j+1}}{N} < k < T_{j+1} \quad (\text{数3})$$

の範囲を満足する時間の範囲kの中の音響信号を削除する。この削除した部分を詰め、残りの部分を順次つなぎ合わせるにより時間軸上の圧縮を行い、実質的に再生速度の向上、すなわち早送り、を行う。伝送系を含めて考えた場合、破棄の操作は伝送前に実行しておくことが効率的である。

【0025】ここで、実際に伝送される係数は時間領域信号系列{b_i}ではなく、周波数領域系列{a_i}なので周波数領域系列{a_i}において上述の(数3)式におけるkに相当する範囲(今後これをmとする)を削除して伝送する。換言すれば、上記(数3)式における時間領域(周波数分解能は有していない。)での範囲kはQMFを用いたMDCTで関係付けられている。時間周波数領域における、このkに等価な範囲をmとして、このmに相当する範囲を破棄して伝送することである。この{a_i}は周波数領域での信号系列を示すものであるが、MDCTの窓長と同じ時間分解能を持つので、時刻の系列{T_j}を求めるとき、その値を該当フレームにおける時間分解能の単位で決定すれば周波数領域における時間長mが求められる。伝送後は従来通りのデコードを行うが、上記削除の処理により無音状態となった時間を詰めて再生する。

【0026】次に、任意のグループのみ再生する方法について説明する。たとえばforegroundのグループを再生する場合、必要となる情報はグルーピング情報{g_i}とg_i=1となるときの信号f_iだけである。伝送系を含めて本発明を実施する場合には無駄な伝送を避けるために、この情報だけを伝送する。上記の2つの情報から上述の周波数領域での信号系列{a_i}を生成し通常の復号手順を適用することで、foregroundのグループ(例えば主旋律)を再生する。以上述べた再生過程においても、上記復号の処理アルゴリズムにしたがったプログラムを予め記録媒体に記録しておき、この記録媒体を使用して上記の復号化処理を行うことにより、上記の復号・再生処理を容易に実行することが出来る。

【0027】

して、時間分離情報(T-分類情報)2gを用いて音響信号の一部削除を行い、残りの音響信号を順次つなぎ合わせる。たとえば、上述の音響信号の系列{b_i}に対して時刻の系列{T_j}が得られたとき、N倍速で再生するためには

【数3】

ら編集加工等が容易な符号系列を生成することができ、符号系列から原音に近い音を再現することができる。

【0028】また、構造化された符号形態を利用することによって、音響信号の早送り再生や、データ削減された音響信号の一部を再生することが容易にできるだけなく、伝送経路を含めての利用を考慮した場合、効率的な伝送を行うことができる。

【図面の簡単な説明】

【図1】本発明における符号化処理系の構成図。

【図2】本発明における復号化処理系の構成図。

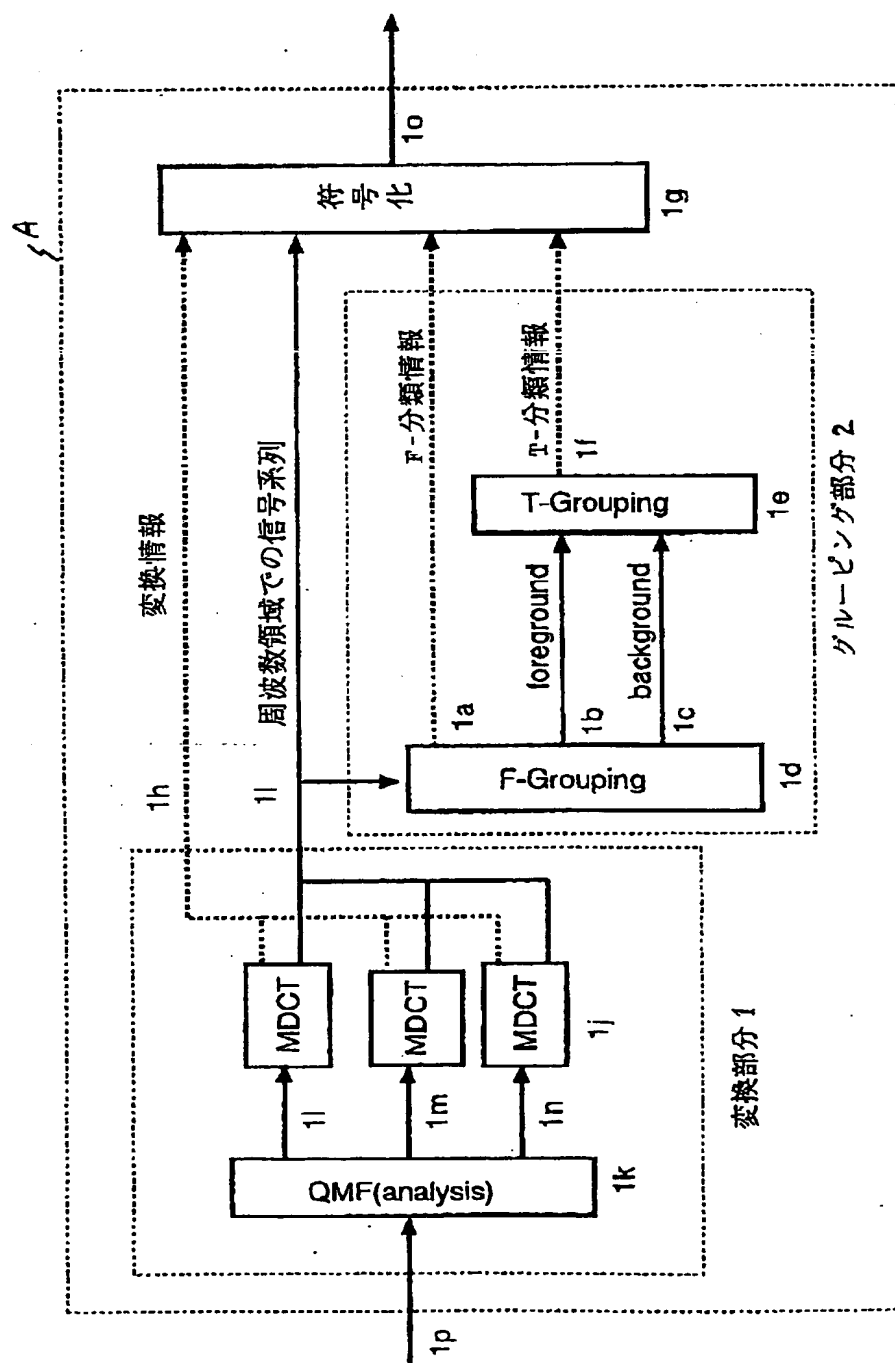
【図3】符号化処理により符号化されたデータ系列を示すフォーマット図。

【図4】時間分離情報のビットストリーム図。

【符号の説明】

1 : 変換部分	2 : グルーピング部
1a : F-分類情報	1b : foreground
1c : background	1d : F-Grouping
1e : T-Grouping	1g : 符号化モジュール
1h : 変換情報	1i : 周波数領域での信号系列
1j : MDCT (analysis)	1k : QMF
1p : 入力音響信号	1o : 符号化出力ビットストリーム
2a : ビットストリーム分解	2b : グループ選択、編集加工
2c : IMDCT (synthesis)	2d : QMF
2e : 周波数領域での信号系列	2f : MDCT変換情報
2g : T-分類情報	2h : F-分類情報
2o : 複合化出力信号	diff : 変数

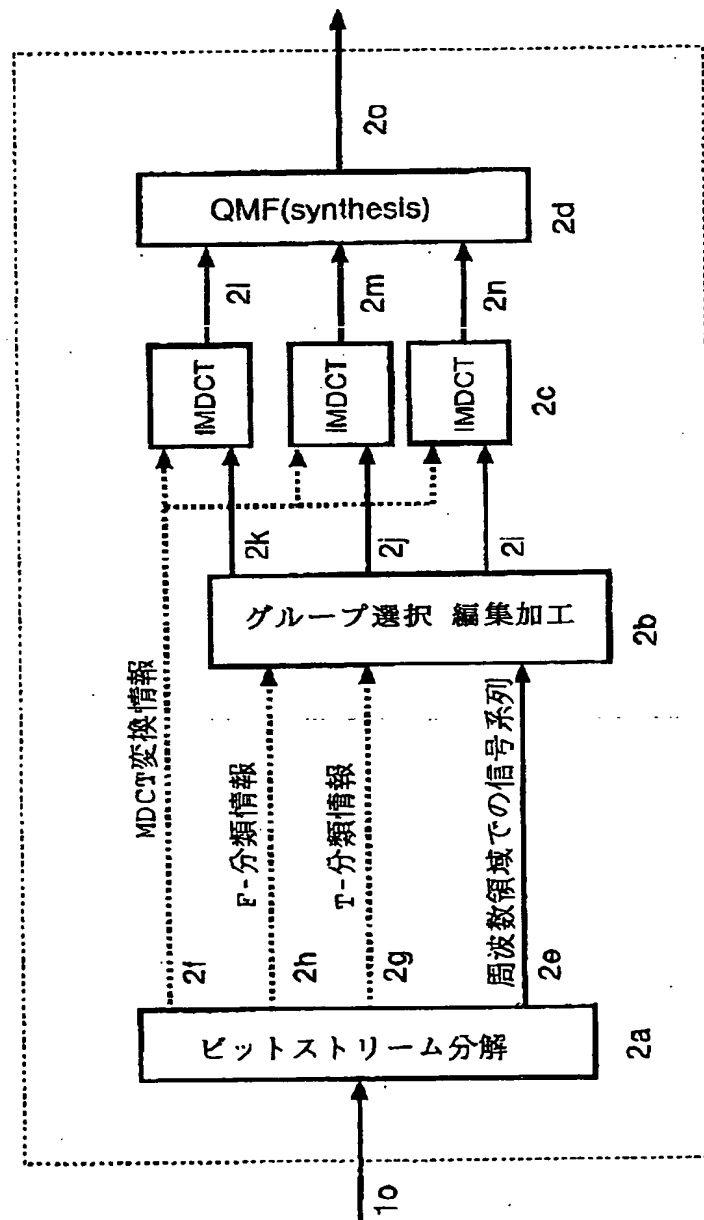
図 1



【図 1】

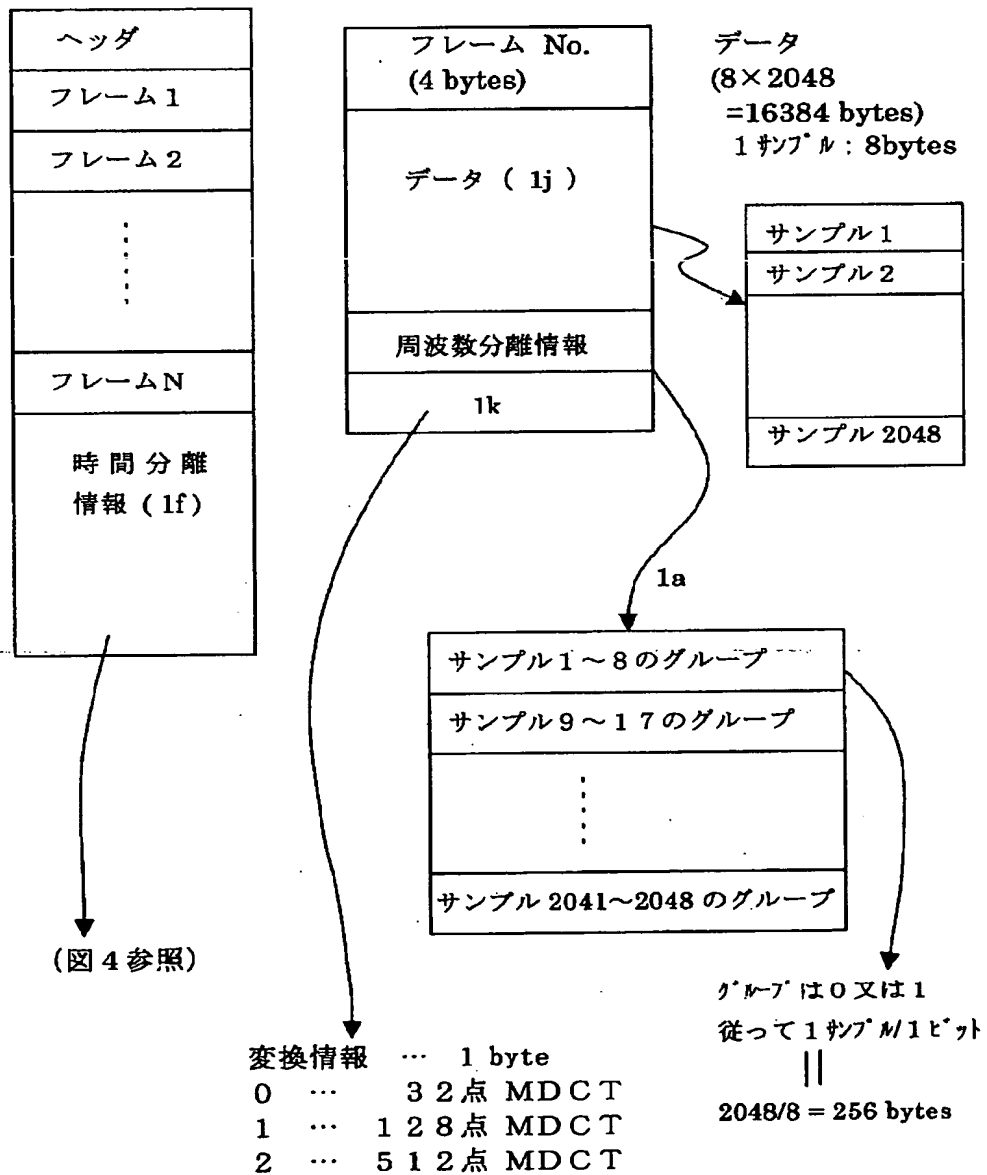
【図2】

図 2



【図3】

図 3



【図 4】

図 4

時間分離情報

